Do Agents Place Too Much Weight on Recent Information?

Or – more particularly – do general managers?

Jane Leber Herr

October, 2005

PRELIMINARY AND INCOMPLETE
Please do not cite without permission

Using data from Major League Baseball's free agent market, I test whether teams place too much weight on recent information in choosing their willingness to pay for a player. In particular, comparing pre- and post-free agency playing statistics, I consider whether the link between free agency salary and previous performance is overly weighted towards recent results. I also test whether the strength of this relationship varies by whether the most recent year's performance represents a better- or worse-than expected outcome. I find that teams irrationally overvalue recent information, but only in response to bad news.

1. Do Agents Overweight Recent Information?

When predicting the future productivity of investment goods, do buyers place too much weight on recent information? Neoclassical economic theory assumes rationality in such market decisions – in particular, that agents will use all available, relevant information. This does not imply that all will place the same value on a given good – there may be valid differences across individuals in the marginal benefit associated with the good or in their risk preferences. Economic theory predicts, however, that buyers will on average correctly estimate their personal return, as based on *all* the information available at the time of purchase.

But do buyers in fact use this information rationally? For instance, in the stock market, might investors place more weight on the recent success of a company than on its long-run performance? If the former is in fact a better predictor of future outcomes, then this finding does not imply irrationality. Such behavior is only inconsistent if buyers systematically place *too much* weight on recent results. Alternatively, buyers may overvalue recent information only if these results themselves varied from expectations. In this latter case, one can then ask if the level of observed 'irrationality' depends on the sign of the deviation.

In this paper I test these hypotheses in Major League Baseball's free agent market. There are two key benefits in using baseball as a test case. First, unlike most employment settings, performance is almost fully observable. If a team makes salary offers as a function of its expectations of future player performance, and if these expectations are based on past performance, then we, too, see the information used in this decision process.

Second, in baseball there exists an open market only for those players who have completed at least six years in the major leagues, labeled 'free agents' because they alone may contract with any team of their choice.[1] Thus, by construction, at the point of free agency eligibility, teams – or more particularly their general managers (GMs) – will have at least a full six years of information available for predicting future performance.

Given this setting and a sample of players who have completed at least one year beyond their free agency eligibility, I can test whether GMs place too much weight on recent information in their salary decisions. In particular, using pre-free agency performance and post-free agency performance and salary, I can compare how the latter two vary with the former. If the relative weight placed on the most recent information is heavier in the *salary* than *performance* equation, this suggests that teams place too much value on recent information. Similarly I can ask whether GMs only overvalue recent performance information when that observed performance itself varied from expectations. To get at these questions I begin below by discussing existing evidence on economic agents' misuse of information. Section 3 will then introduce my methodology, Section 4 my data, and Section 5 my results. As you will see I find no clear evidence that agents overvalue recent information in general, but clear evidence that they overvalue this information when it itself was below expectations. Section 6 concludes accordingly.

2. Existing Economic Studies on the Misuse of Information

- Rabin (1999) – *First Impressions Matter: A Model of Confirmatory Bias*, **QJE**, Feb 99 (this paper implies that agents overweight earliest information).
- Grether (1992) – *Testing Bayes Rule as a Descriptive Model of The Representativeness Heuristic*, **QJE**, 95 (537-57).
- Ouwersloot, Nijkamp & Rietveld (1998) – *Errors in Probability Updating Behaviour: Measurement and Impact Analysis*, **Journal of Economic Psychology**, 19 (535-63).

[1] In a historic legal fluke Major League Baseball is exempt from monopoly regulations. Under the current rules all players in the minor leagues, as well as those in their first six years in the majors, may only contract with their present team. (A year is defined as 172 days on the major league team's 25-man roster or on injured reserve.) Note that the market for players clearly exists during the pre-free agency years, it simply exists between teams rather than between the team and player.

- Zizzo, Stolarz-Fantino, Wen & Fantino (2000) – *A Violation of the Monotonicity Axiom: Experimental Evidence on the Conjunction Fallacy*, **Journal of Economic Behavior and Organization**, 41 (263-276).

3. Methodology

In this paper I ask whether baseball general managers place too much weight on recent player performance, relative to the full information set available, when determining their willingness to pay for free agents. Let us assume that GMs place value on players as a function of their expected capacity to increase the team's winning percentage.[2] If player *i*'s contribution to wins is a direct function of his on-field performance, we should expect GM's willingness to pay for free agent *i*, y_i, to be a function of his expectation of player *i*'s future performance, X_i: $y_i = f(X_i)$.

The question is how GMs generate their expectations. Suppose the true relationship is some function, $X_{i7} = g(x_{ik})$, for $k = 1$ to 6, where we focus on the outcome year 7 performance, and x_{ik} represent *i*'s pre-free agency annual performance. But suppose that GMs misperceive expected performance as $g'(x_{ik})$, in which g' places too much weight on most recent information, in this case year 6. For instance, suppose g is a linear function of past performance:

$$g(x_{ik}) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \varepsilon_i \tag{1}$$

but that GMs overestimate β_6 such that

$$g'(x_{ik}) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6' x_{i6} + \varepsilon_i \tag{1'}$$

with $\beta_6' > \beta_6$.

---

[2] In truth we expect GMs to value players as a function of their ability to increase team *profits*, driven by ticket sales, TV contracts, merchandise, etc. Although profits are highly correlated with wins, some players will generate revenue simply by their 'star' power. I will attempt to control for this by including information on All Star game appearances.

To test for this, taking a sample of players who have completed year 7, we can compare how observed post-free agency performance and salary vary with pre-free agency performance. If *g* is the true data generating process, then we should observe post-free agency performance X_{i7} following Equation (1). In contrast, observed salary will be a function of g', $y_i = f(g'(x_{ij}))$. Again making the simplifying assumption that *f* is a linear transformation, the salary function expands to

$$y_i = \gamma\,(\alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6{}' x_{i6} + \varepsilon_i\,)$$

$$= \alpha' + \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3} + \theta_4 x_{i4} + \theta_5 x_{i5} + \theta_6 x_{i6} + \varepsilon_i{}' \qquad (2)$$

where $\theta_j = \gamma\beta_j$ for j = 1 to 5, and $\theta_6 = \gamma\beta_6{}'$.[3] Thus if GMs place too much weight on recent information, then the ratio of the coefficient on year 6 to previous years' performance in Equation (2) should be greater than the comparable ratio in Equation (1):

$$\theta_6/\theta_j \;=\; \gamma\,(\beta_6{}'/\beta_j\,) \quad > \; \beta_6/\beta_j$$

An alternative way of thinking about this is to rewrite $\beta_6{}'$ as $\beta_6{}' = \beta_6 + \delta$, with $\delta > 0$. Thus Equation (2) becomes

$$y_i \qquad = \gamma(\beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + (\beta_6 + \delta) x_{i6} + \varepsilon_i\,)$$

$$= \gamma(\,g(x_{ik})) + \gamma\delta x_{\,i6} \qquad (2')$$

where the first term in Equation (2') is a function of the true expected value of X_{i7}. I can therefore use my sample to fit the observed post- (year 7) and pre-free agency performance results to build

---

[3] Alternatively I can assume that GMs expectations of future player performance are rational ($g = g'$), and that it is in this second step that performance in year 6 is overweighted ($\gamma_6 > \gamma_5$, γ_4, etc.). This would imply that GMs believe that year 6 performance has an especially large value in determining a player's marginal revenue product in their post-free agency years. To me this seems less reasonable than assuming that GMs incorrectly believe that year 6 performance has a higher relative capacity to predict future performance.

an in-sample estimate of $X_{i7} = g(x_{ik})$. Regressing post-free agency salary on both this measure and year 6 performance, a significant positive coefficient on the latter will provide additional evidence that GMs overweight recent information in making salary decisions.

The second question posed is whether GMs only place too much weight on recent information when these results themselves were unexpected. Suppose, similar to Equation (1), there exists a true data generating process for year 6 performance, $x_{i6} = h(x_{ij})$, for $j = 1$ to 5. Suppose now that in place of Equation (1′), GMs have an alternate irrational expectations function:

$$g''(x_{ij}) = \alpha + \beta_j x_{ij} + \varphi_{6+}[1(x_{i6} > h(x_{ij}) + U_+)] + \varphi_{6-}[1(x_{i6} < h(x_{ij}) - U_-)] + \varepsilon_i \qquad (1'')$$

again for j = 1 to 5. Thus g'' is dependent on whether year 6 performance fell above or below some expected level (defined by upper and lower boundaries U_+ and U_-). Assuming this alternate expectations function, we will observe the following salary function:

$$y_i = \gamma(\alpha + \beta_j x_{ij} + \varphi_{6+}[1(x_{i6} > h(x_{ij}) + U_+)] + \varphi_{6-}[1(x_{i6} < h(x_{ij}) - U_-)] \qquad (2'')$$

Testing this second hypothesis becomes more difficult because I do not know the values of U_+ and U_-. As a rough proxy, assuming h is linear in previous performance with normally distributed errors υ_i, I use my sample to estimate the per-player residual: $\upsilon_{i\_hat} = x_{i6} - x_{i6\_hat}$. I then define U_+ and U_- as a fraction of the sample standard deviation of these errors:

$$U_+ , U_- = +/- C^*\text{s.d.}(\upsilon_{i\_hat})$$

Using these thresholds I define two dummy variables to capture the population of players whose 6^{th} year performance was greater or less than their predicted performance plus the relevant threshold. Following Equation (2″), if the regression coefficients on either are significantly different from zero, this implies that players receive an (irrational) salary bonus or penalty if their most recent performance varied from expectations. If the first hypothesis – that GMs overvalue recent information in general – cannot be rejected, it may be that recent information is only used irrationally under these circumstances.

4. Data

Unlike most employment settings, in baseball there is a vast amount of worker-specific information readily available. For this analysis I rely on the Lahman Baseball Database, which includes an astounding wealth of player-specific information from 1871 to the present.[4] For my analysis I use information on salaries, offensive and fielding statistics, and All Star appearances for those players debuting from 1978 forward.

Contract Details

One piece of information not easily available, however, is contract signing and start dates. This lack of contract information complicates my analysis because not all players sign new contracts after their sixth full year in the majors. Thus when comparing year 7 performance and salary to earlier data, year 6 may fall *after* year 7 salary was set.[5] Among this subset, year 4 or year 5 – rather than year 6 – will be the 'latest' information at the time of the salary decision. If GMs truly place too much weight on recent information then it will be the coefficients on these *earlier* years that are over-weighted in the salary equation. Since my first test is to compare the coefficient ratios of year 6 to earlier years across the salary and performance equations, if the earlier coefficients are inflated via this same mechanism for this subset of players for whom year 7 is not a new contract, my measure will tend to be an underestimate of the true ratio.[6]

---

[4] Version 5.2. The data can be downloaded from The Baseball Archive, www.baseball1.com.

[5] Even for players starting a new contract in year 7, among those who sign with their original team, some will have signed before their sixth season was played. Thus the same complication holds.

[6] A second potential complication is that the subset who do not sign a new contract in year 7 may not be a representative sample. For instance, teams my offer contracts that extend beyond year 6 to their better young players. For this to drive my results, however, there would need to be a systematic spuriously large

As one means of addressing this complication, for those players on the same team in years 6 and 7 I will also run the analysis on year 8 salary, with most recent information now defined as year 7. Although some such players will in fact be starting a new contract (signed with their original team), and some players switching teams will be continuing on an old one (they were traded), this split is my best means at proxying for those signing a new contract after year 7. But to the extent that some within this subset sign a new, multiyear contract in year 7 – and thus year 7 statistics occur *after* the year 8 salary decision is made – the same complication will hold.

Sample and Variable Specification

Because the Lahman database only includes salaries starting in 1985, I am forced to limit my analysis to players who debuted from 1978 forward. For ease of analysis I consider only batters, and focus my attention on the importance of their offensive statistics in predicting future salary. From 1978 through 2004, 1,418 players played in at least seven years in the major leagues. This calculation, however, captures anyone who played at least one game in seven different seasons; it does not distinguish how many completed the full six years' service time necessary to achieve free agency.

Unfortunately the Lahman database does not list a player's service time – it instead includes data only for those games in which he played. The only relevant information available is a player's debut month; players debuting after the start of the season clearly cannot accrue a full year's service during their rookie year.[7] To estimate the year of eligibility (all contracts begin

correlation between year 6 performance and year 7 salary, even though year 7 salary was set before year 6 was played.

[7] For 1998 I lack all data on debut month. Given that the mean number of ABs for players debuting in April in the remaining years is 119 (compared to 46 for those debuting in later months). I therefore use this (rather arbitrary) cut-off for defining whether 1998 is year 1 for those debuting that year. The Lahman database also lacks information on how often players move back and forth between the majors and minors, as is common for young players. Thus even players debuting in April may accrue less than a full season of service time that year. However my sense is that rookies who begin the season on the major league team will tend to be better and more experienced, and thus are less likely to move back and forth than those who come up midseason.

with the start of a season) I therefore define a player's first year as year 1 only if he debuts in April, the start of the season. Using this definition, only 1,213 players reached their 7th full year of play in the majors.

Up till this point I have only dropped players who clearly never reached free agency. In the following step, however, I limit my sample to those with at least 100 plate appearances in each of their first seven years.[8] Although this decreases my sample size appreciably to 320 players, this focuses my analysis on those players for whom general managers will have the most complete information. If we think that agents are least likely to misuse information when faced with a complete information set, any evidence of the overvaluation of recent information among this group implies that the true effect holds regardless of the completeness of the information available.[9] (If I instead allow one year of fewer than 100 plate appearances – in all but the sixth year – I end with a final sample of 443 players compared to the final sample of 302 discussed below. Running the analysis on these gives substantially the same results as those presented below.) Comparing average plate appearances for those players remaining in my sample versus those excluded on this criteria, the former have an average 478 per year during their first seven years versus only 114 for the latter.[10]

Lastly I limit my sample to those players with seven consecutive years played. Some players may miss a full year because of a serious injury. Although they may continue accruing service time, this period does not provide any information by which to gauge predicted future

[8] The Lahman database does not include information on total plate appearances. As discussed in footnote 13, my proxy measure will be a slight underestimate of true plate appearances per year. (For those with their debut year defined as year 0, I do not require 100 plate appearances in that year.)

[9] One might worry that the players observed to reach year 7 are a select group – that those who find themselves with a low free-agency wage offer may choose to leave baseball. Yet because major league baseball salaries are so much higher than those likely to be available in the regular labor force to a marginal player (with maybe a college education?), I think this worry is unnecessary.

[10] Assuming roughly four plate appearances per game, for the former this amounts to approximately 120 of the 162 games in the season, versus only 29 for the latter.

performance.[11] This requirement limits my sample to 309 players. By necessity, I am also limited to those with salary information for year 7. This provides a final sample size of 300. Of this 300, 200 play on the same team in years 6 and 7. For this subset, to use year 8 as my left-hand side variable I again require the minimum number of plate appearances and consecutive years played. Thus of this 200, 172 (86 percent) have sufficient data

Variable Specification

To measure offensive performance I create a variable, slugging-on-base percentage (SOB), which is a weighted average of the proportion of times a player gets on base, with the weights representing the corresponding number of bases:[12]

SOB = [1*(1B + BB + IBB+HBP) + 2*(2B) + 3*(3B) + 4*(HR)]/PA

where:
- 1B = single
- 2B = double
- 3B = triple
- HR = homerun
- BB = base on balls (a 'walk')
- IBB = intentional base on balls
- HBP = hit by pitch
- PA = plate appearance

[11] Furthermore the injury itself will likely play into a GM's prediction of a player's future performance, or simply his longevity. Ideally I would like to include a control for injuries, but I lack such data. Excluding players with missed years is one way of culling out the more egregious cases.

[12] For those unfamiliar with the game let me provide a quick primer. The baseball diamond is made up of four 'bases'. When a player comes to bat, if he can get on base and make it around all four before three 'outs' are made, he scores a run. To get on base he must either get a hit, be awarded first base (through a walk or by being hit by a pitch), or via an error by the opposing team. A hit can be either a single (reaching only as far as first base), a double (reaching second), a triple (reaching third) or a homerun (reaching home

This percentage therefore gives a weighted representation of how often a player gets on base as a function of all such opportunities.

Those who follow baseball will not recognize this statistic. The more traditional measures are batting average (BA), slugging percentage (SB), and on-base percentage (OBP). The former is the proportion of 'at-bats' (AB) in which a player makes any type of hit:

BA = [1B + 2B + 3B + HR]/AB

Note that the denominator in this fraction is only a subset of all plate appearances; 'at bats' subtract all instances in which a player gets on base without a hit (for instance via a walk), or in which he hits a sacrifice (a play that advances a base runner but results in an out).[13]

Slugging percentage begins with the structure of the batting average – including only hits, and as a percentage of at-bats – but weights each hit by the corresponding number of bases achieved:

Slugging = [1*1B + 2*2B + 3*3B + 4*HR]/AB

In contrast, on-base percentage is exactly that – the proportion of plate appearances in which a player gets on base, regardless of how he does so:

OBP = [1B + 2B + 3B + HR + BB + IBB + HBP]/PA

A more recently discussed statistic, 'on-base plus slugging' (OPS), is simply the sum of a player's slugging and on-base percentages. I consider this a cruder version of my SOB statistic,

---

base). A 'base on balls', otherwise known as a 'walk', occurs when a pitcher throws four 'balls' (pitches thrown outside of the strike zone) before throwing three 'strikes'.

[13] The at-bat is so ingrained as the traditional measure that the Lahman database does not list plate appearances. As a proxy I sum at-bats, walks, sacrifices, and instances when a player is hit by a pitch. This is in truth an underestimate of total plate appearances because it does not capture instances in which players get on base via a defensive error of the opposing team.

but in case OPS is the measure that GMs consider in making their pricing decisions, I run my analysis on both, as well as on batting average alone.

Because salary decisions are not driven solely on offensive skill, I include in both the performance and salary equations the following control variables that capture other characteristics valued by GMs:[14] defensive position in year 6, average fielding percentage in years 4 through 6, and the number of All Star appearances during years 1 through 6. For the first, because some positions have a heavier defensive load, all else equal teams may be willing to pay more for those who play one position (e.g., shortstop) than another (outfield).[15] The second and third factors similarly may drive salary offers. Fielding measures defensive ability[16] and (as discussed in footnote 2**)**, All Star appearances may capture elements of 'star power' that can raise a team's profits through avenues other than wins.

I also control for whether a player was born in the same state as his post-free agency team, and whether he played with that same team pre-free agency. Because a player may be willing to accept a lower salary in order to play in his home state, I would like to include a control for whether his post-free agency team is located in the same state in which he lives during the off-season.[17] Unfortunately only state of birth, not residence, is available in the Lahman database. Similarly, a player may be willing to accept a lower contract in order to stay with his existing

[14] I expect most to only matter in the salary equation, but I include them in both for consistency.
[15] For those unfamiliar with baseball, there are nine defensive positions, broken into four general types: the pitcher, catcher (behind home base), infielders (those who play in the region surrounding the bases), and outfielders (those who play in the 'outfield'). Infielders include the first-, second- and third-basemen, and shortstop. For my analysis I create 7 position dummies: catcher, first through third base, shortstop, outfielder, and utility infielder. (A 'utility infielder' is a player with no regularly-assigned position, generally back-up players. Because many players will play more than one position over the course of the season, I define a player's position as that which he played in at least 50 percent of his year 6 games.
[16] I create a measure of the proportion of a player's defensive plays during years 4 through 6 in which he does *not* commit an error: Fielding = 1 – (E/(A+PO)), where E = error, A = assist and PO = put outs. (An error is an event in which a player makes a defensive mistake that benefits the opposing team.)
[17] Because the baseball season lasts only 6 months and trades occur often, many have their official residence (where their families live year-round) in states other than the state in which they play.

team if this minimizes family disruption or if he has developed a strong bond with his original team.

Lastly I control for age and team and year fixed effects for the post-free agency year.[18] Considering team fixed effects, a player's marginal revenue product is a function not only of the player's skill, but also of the team's relative market size and penetration. The New York Yankees, with a market of more than 10 million, will generate more revenue from any given player than will the Kansas City Royals, and thus should be willing to pay more.

5. Results

To begin, Table 1 lists sample averages for the dependant variables – free-agency salary and performance – as well as the most recent pre-free agency performance and the controls just discussed. The middle column lists these values for the sample as a whole (when using year 7 as the post-free agency year), while the right-hand column lists these values for the subset of players for whom I can use year 8 as my first post-free agency year, namely those who played for the same team in years 6 and 7.[19]

The first panel of Table 1 lists average post-free agency salaries and performance levels, including all three statistics discussed above. (Notice that these players get paid a lot of money.) The next panel lists these same performance measures for the most recent pre-free agency year. As you can see, performance in the most recent year is almost identical to that in the first post-free agency year. Hence on average, general managers should have little difficulty predicting future player performance.

[18] Because players can be traded midseason, for many I have more than one team per year. For the purposes of this variable I assume that the first team played for in that year is the team making the salary decision.
[19] As discussed above, although this will capture a subset of players who simply signed with their original team, it will also capture those players who were signed through their seventh year under a pre-free agency contract.

Table 1. Descriptive Statistics				
	Year 7 as LHS		Year 8 as LHS[1]	
	Mean	**(s.d.)**	**Mean**	**(s.d.)**
Post Free Agency:				
Salary (2004$)	$ 3,647,004	$ 3,057,064	$ 4,095,906	$ 3,072,928
Slugging-on-base percentage (SOB)	0.486	0.089	0.490	0.083
On-base plus Slugging (OPS)	0.778	0.127	0.783	0.119
Batting average	0.271	0.035	0.273	0.034
Most Recent Pre-Contract Year:				
Slugging-on-base percentage (SOB)	0.488	0.083	0.496	0.091
On-base plus Slugging (OPS)	0.781	0.116	0.793	0.131
Batting average	0.272	0.031	0.276	0.036
Controls:				
Post-free agency team located in state of birth	0.10	0.30	0.13	0.33
Playing for same team pre- and post-free agency	0.67	0.47	0.66	0.48
Total pre-free agency All Star appearances	0.97	1.40	1.41	1.78
Age in Free Agency Year	30.1	1.64	30.9	1.7
Position:				
First Base	47		27	
Second Base	32		25	
Third Base	32		18	
Short Stop	31		18	
Catcher	33		20	
Outfield & Designated Hitter	120		62	
Utility Infielder	7		4	
Sample Size:	302		174	

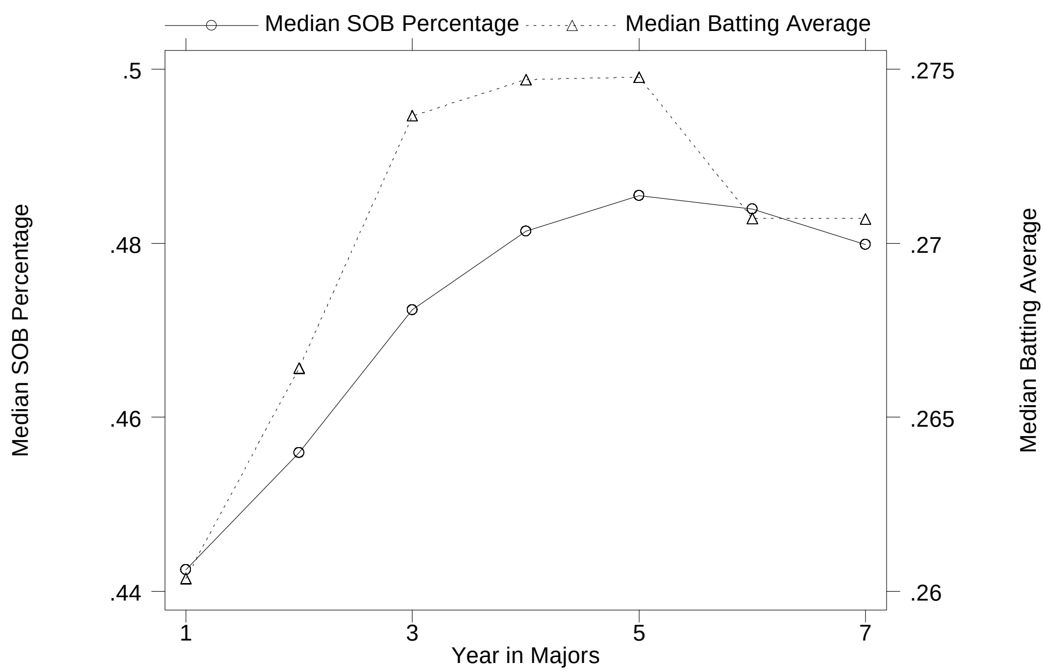
NOTES:

(1) Since some players observed to be playing on the same team in years 6 and 7 may not begin a new contract till year 8, I rerun the analysis on this subsample using Year 8 and my post-free agency year and year 7 as the 'most recent' information. See discussion in text.

The final panel lists several of the controls included in the analysis. One statistic to note is the proportion of players staying with the same team: 67 percent when considering year 7 as the first post-free agency year, or 66 percent if considering year 8. This first percentage, taken alone, might indicate that a large number of players do not sign a new contract before year 7. Yet unless many sign contracts that extend through year 8, this second percentage appears to belie this interpretation.[20] If instead it is simply that most players sign with their original team, then year6, not year 7, will be the most recent information at the contract signing date, and using year 8 as my left-hand side variable to year 7 as my right-hand side should give me much less clear results.

Before I go on to the econometrics, let us first consider performance trends. Assuming that general managers form their expectations of future player performance (their prior) based on past trends, we can use this sample to consider how offensive performance varies over time. Using my sample, Figure 1 shows the median performance level per year for these players' first 7 years in the majors. As a comparison, the right-hand margin shows median slugging-on-base (SOB) percentages and the left-hand margin shows batting averages.[21]

Figure 1. Performance Over Time



[20] Among my 300 players, 75 percent are on the same team in years 5 and 6, versus 83 to 90 percent when considering team retention from years 1 through 5.

[21] A comparable graph including OPS percentage shows the same pattern, although with more movement in the last 4 years. In order for SOB to remain high as the batting average falls in year 6, players on average

We see that both increase quickly through the first three to four years and then roughly level off, although the growth over time in median performance is only between 5 and 10 percent. As we could predict given our results in Table 1, performance in year 6 is very close to that in year 7. In fact, this graph shows that, on average, players achieve a roughly steady state performance level by year 3 or 4. Table 2 shows this same result: considering the correlation between year 7 performance and earlier years, for any of the three statistics we see a distinct increase in correlation between years 3 and 4.

Table 2: Offensive Performance Correlations of Year 7 With Previous Years

	Year 6	Year 5	Year 4	Year 3	Year 2	Year 1
Slugging on-base percentage (SOB)	0.679	0.643	0.634	0.479	0.491	0.444
On-base percentage plus slugging (OPS)	0.623	0.587	0.587	0.399	0.438	0.357
Batting average (BA)	0.401	0.401	0.394	0.197	0.315	0.148

As discussed above in Section 3, one means for testing whether GMs place too much weight on recent information is to compare across the post-free agency performance and salary equations the coefficient ratio of the most recent year to earlier years. If agents overvalue recent information in their pricing decision, we will expect the ratios to be larger in the salary than in the performance equation.

Table 3a lists the results of the linear regression of the log of year 7 performance and salary on the log of previous performance and the controls listed above. (Because of the pattern observed in Figure 1, I limit my discussion on previous years' data to years 4 through 6.) Looking at the performance regression, as expected we see that the last three years' results are all highly

must be getting more extra-base hits or getting on base at a higher rate. Exploring the data it is clear that the latter is the more important factor.

correlated with year 7, even despite the high level of multicollinearity. Of the controls, only year fixed effects are significant.

Table 3a - Comparison of Post-Free Agency Performance and Salary Coefficients (year 7)

Dependent Variable:	**ln(Performance$_7$)**			**ln(Salary$_7$)**		
	Coefficient	**s.e.**	**p-value[1]**	**Coefficient**	**s.e.**	**p-value[1]**
$\ln(SOB_6)^2$	0.325	0.071	**0.000**	1.024	0.310	**0.001**
$\ln(SOB_5)^2$	0.181	0.078	**0.022**	0.525	0.342	0.126
$\ln(SOB_4)^2$	0.380	0.082	**0.000**	0.713	0.357	**0.047**
Coefficient Ratios:						
(year 6)/(year 5)	**1.80**		-	**1.95**		-
(year 6)/(year 4)	**0.86**		-	**1.44**		-
Age	-0.008	0.005	0.124	-0.071	0.023	**0.003**
Total All Star Appearances (yrs 1-6)	-0.001	0.007	0.827	0.127	0.029	**0.000**
Position (omitted = outfield)	-		0.275	-		**0.040**
1st Base	0.024	0.025	0.340	-0.113	0.107	0.293
2nd Base	-0.050	0.028	0.073	0.037	0.122	0.761
3rd Base	0.012	0.032	0.702	0.150	0.139	0.283
Shortstop	-0.017	0.031	0.588	0.332	0.137	**0.016**
Catcher	-0.033	0.028	0.247	-0.196	0.124	0.114
Fielding Percentage (average years 4-6)	0.645	0.823	0.434	9.402	3.593	0.296
Team Fixed Effects (year 7)	-		0.173	-		0.244
Same Team Years 6 and 7	-0.007	0.017	0.670	0.382	0.075	**0.000**
Year 7 team located in state of birth	0.025	0.028	0.368	0.053	0.120	0.657
Year 7 Fixed Effects			**0.003**	-		**0.000**
Adjusted R^2:	**0.57**			**0.64**		
Sample Size:	**302**			**302**		

NOTES:

(1) For dummy variables (e.g., position or team), the listed value is the joint significance.

(2) $\ln(SOB_t)$ = Log of slugging-on-base percentage for year t.

As expected, in the salary equation many more of the controls are significant: older players are paid less, each trip to the All Star game increases wages (even holding constant performance), and defense matters. One potentially surprising result is the large and significantly positive coefficient on whether one played for the same team in years 6 and 7. Remember that I predicted a negative coefficient – that players might accept lower contracts in order to minimize the family disruption of moving. Yet this coefficient instead suggests that relatively better players resign with their original team.[22]

Returning to the primary question at hand, in comparison to the performance regression, in the salary regression the p values for previous performance are all appreciably higher for all but the most recent year. Comparing the corresponding coefficient ratios, we do find that they are larger in the salary equation, although not appreciably so. Rerunning this analysis with the larger sample of 443 players gives the same results. As a second comparison I run this analysis using OPS (Appendix Table 1) and batting average (Appendix Table 2), in case these are instead the performance measures considered by general managers. In all four cases the coefficient ratios are larger in the salary equation, but not by much.

Rerunning this analysis using year 8 as my measure of the first post-free agency year (see Table 3b), I find less clear results. In either regression the partial correlations with earlier years' performance are generally lower and less significant, and the corresponding coefficient ratios are less intuitive. Using the larger sample (251 players when allowing one year with fewer than 100 plate appearances), the coefficient ratios are less variable.[23] Some of this difference may derive from the smaller sample size – rerunning this with the larger sample you do observe smaller p-values for previous years' performance. Yet much of this may be driven by the miss-specification of year 7 as a pre-free agency year for many of the players in this subsample. Thus in sum, neither Tables 3a nor 3b strongly suggest that agents overvalue recent information.

[22] It may be that relatively worse players receive better offers from other teams who will likely have less information about any potentially unobservable factors of performance than the player's starting team.

[23] In particular, the ratios corresponding to those shown in Table 3b are 1.47, 1.55 and 1.77 in the performance regression (the first represents the ratio of the coefficient on year 7 to year 6), and 2.96, 0.79 and 2.16 in the salary regression.

Table 3b - Comparison of Post-Free Agency Performance and Salary Coefficients (year 8)						
Dependent Variable:	**ln(Performance$_8$)**			**ln(Salary$_8$)**		
	Coefficient	**s.e.**	**p-value[1]**	**Coefficient**	**s.e.**	**p-value[1]**
ln(SOB$_7$)[2]	0.305	0.096	**0.003**	0.872	0.391	**0.028**
ln(SOB$_6$)[2]	0.190	0.111	0.083	-0.077	0.453	0.866
ln(SOB5)[2]	0.041	0.127	0.785	1.432	0.519	**0.007**
ln(SOB$_4$)[2]	0.117	0.120	0.340	0.083	0.492	0.867
Coefficient Ratios:						
(year 7)/(year 6)	**1.61**		-	**-11.35**		-
(year 7)/(year 5)	**7.53**		-	**0.61**		-
(year 7)/(year 4)	**2.61**			**10.57**		
Age	-0.025	0.008	**0.002**	-0.153	0.032	**0.000**
Total All Star Appearances (yrs 1-6)	-0.009	0.008	0.218	0.058	0.032	0.072
Position (omitted = outfield)	-		0.275	-		0.159
1st Base	-0.052	0.036	0.130	0.024	0.145	0.869
2nd Base	-0.023	0.036	0.545	0.032	0.146	0.826
3rd Base	-0.025	0.049	0.599	0.120	0.199	0.549
Shortstop	-0.067	0.049	0.188	0.183	0.199	0.358
Catcher	0.010	0.040	0.896	-0.334	0.163	0.043
Fielding Percentage (average years 5-7)	-0.720	1.384	0.752	0.376	5.657	0.947
Team Fixed Effects (year 8)	-		0.979	-		0.057
Same Teams Years 7 and 8	0.011	0.026	0.710	0.296	0.106	**0.006**
Year 8 team located in state of birth	0.032	0.040	0.396	-0.099	0.165	0.551
Year 8 Fixed Effects	-		0.263	-		**0.000**
Adjusted R^2:	**0.52**			**0.70**		
Sample Size:	**174**			**174**		

NOTES:

(1) For dummy variables (e.g., position or team), the listed value is the joint significance.

(2) ln(SOB$_t$) = Log of slugging-on-base percentage for year t.

The second test of whether GMs place too much weight on recent information is to generate a predicted post-free agency performance measure, given pre-free agency observables, and regress salary on this predicted value as well as on recent performance alone. A significant coefficient on the latter, given that this information is already incorporated into the former, would imply that GMs place too much weight on this recent information.

Table 4 lists the results of this test using both year 7 and year 8 salary as the left-hand-side variable. As seen here, in either case it does not appear that GMs overweight recent information. Whereas in both versions predicted performance is clearly correlated with the post-free agency salary, the most recent year's data is not separately significant.[24] [Ken, will you check this footnote – in particular is the second sentence correct? If not, do you have an idea how I can get appropriate standard errors?] Given these findings and those from Tables 3a and 3b there is no clear proof that teams place too much weight on recent information in determining their willingness to pay. There remains the question: do they overweight recent information only when that information itself varied from expectations?

Table 4: Second Test of Overweighting of Recent Information

Regressors:	**Coefficient**	**s.e.[1]**	**p-value[1]**	**Coefficient**	**s.e.[1]**	**p-value[1]**
Predicted Performance	2.577	0.709	0.000	2.584	1.126	0.024
Recent Performance	0.058	0.500	0.908	0.000	0.697	1.000
Sample Size	302			174		
Adjusted R^2	0.64			0.70		
(Adjusted R^2 of prediction)	0.52			0.48		

NOTE:

(1) Because the predicted performance regressor is predicted, yet OLS treats it as precise, the standard errors (and hence the p-values) for this regressor will be too small.

Above I proposed testing this second question using the following approach. First, using year 1 through 5 observed results and a linear specification, I create a predicted year 6 performance measure. Comparing this prediction to the observed year 6 performance I then calculate the per-player prediction error, u_{i_hat}. Finally, using the sample distribution of this error, I define threshold levels U_+ and U_-. Those players with observed year 6 results larger than their predicted value plus U_+ are considered to have played 'better than expected', while those with

[24] Note that the standard errors listed for the predicted performance regressor are too small because ordinary least squares treats the regressor as a fixed point when in fact it is only a prediction. This should

year 6 results smaller than their predicted value plus U_ played 'worse than expected'. (Using year 8 as my post-free agency period, I follow the same approach using observed performance in years 1 through 6 to predict year 7 results.)

Table 5 lists the results. The top panel lists the mean and standard deviation of the predicted log of slugging-on-base percentage (SOB) for the most recent year preceding free agency. For year 7 this mean corresponds to a SOB percentage of 0.488, for year 8 it a larger 0.496. This top panel also lists the thresholds associated with each of the tests below. In the first instance, I set the upper and lower bounds, U_+ and U_-, equal to one standard deviation of the sample distribution of u_{i_hat}. Using this definition, with year 7 as my post-free agency year 47 players played better than expected in year 6 and 44 played worse.

I then regress year 7 salary on year 1 through 5 performance, the controls listed above, and the two dummies defining 'beyond expectations'. Note that I do not include year 6 observed or predicted performance. By construction the latter is simply a linear function of year 1 through 5 performance, and hence will be dropped. The former is excluded because, given this implicit inclusion of predicted year 6 performance, including year 6 *observed* performance is mathematically equivalent to including the prediction error, u_{i_hat}. Because my variables of interest are a function of this error term, including this underlying variable will absorb their effect and lead to no clear conclusion.[25]

---

not, however, affect the estimation of the standard error for recent performance.

[25] In particular, if I were to include observed year 6 performance, and hence effectively u_{i_hat}, we will be testing for the effect on salary of playing beyond expectations, holding constant actual observed play. Yet if a player played beyond expectations in year 6, by construction this means that his year 5 and earlier performance was poorer than a player with an equivalent year 6 performance who did *not* play beyond expectations. Thus, all else equal, for these two players we should expect the coefficient on the dummy for playing above expectations to be *negative*. When I run these regressions, I do find just that in two of the four specifications (using the 2 thresholds and the two versions of post-free agency data), and when positive the coefficients are never significant at traditional levels.

Table 5: Coefficients on Beyond Expected Performance Dummies				
	Year 7		**Year 8**	
	Mean	**Standard Deviation**	**Mean**	**Standard Deviation**
ln(SOB)	-0.732	0.125	-0.719	0.142
u_{hat}	4.93×10^{-10}	0.115	-1.20×10^{-9}	0.122
$U_{+/-1}$	+/- 0.115		+/- 0.122	
$U_{+/-1.5}$	+/- 0.172		+/- 0.182	
Threshold: +/- 1 standard deviation				
	Above Threshold	**Below Threshold**	**Above Threshold**	**Below Threshold**
U_{+1}	47	-	30	-
U_{-1}	-	44	-	27
	Coefficient	**(s.e.)**	**Coefficient**	**(s.e.)**
$1(x_{6i} > x_{i6\_hat} + U_{+1})$	0.076	0.099	0.194	0.120
$1(x_{6i} < x_{i6\_hat} + U_{-1})$	-0.243**	0.100	-0.305**	0.130
Threshold: +/- 1.5 standard deviations				
	Above Threshold	**Below Threshold**	**Above Threshold**	**Below Threshold**
$U_{+1.5}$	12	-	8	-
$U_{-1.5}$	-	24	-	16
	Coefficient	**(s.e.)**	**Coefficient**	**(s.e.)**
$1(x_{6i} > x_{i6\_hat} + U_{+1.5})$	0.160	0.179	-0.349	0.223
$1(x_{6i} < x_{i6\_hat} + U_{-1.5})$	-0.381***	0.130	**-0.505*** **	0.166
NOTE: * significant at 10%, ** significant at 5%, *** significant at 1%				

The second panel of Table 5 lists these results. Setting the expectations thresholds equal to one standard deviation of the sample distribution of u_{i_hat}, with year 7 as my post-free agency year I find a positive but insignificant coefficient for players who played better than expected, but a substantially negative coefficient for those who played worse than predicted. The latter face a 25 percent penalty (significant at the 5 percent level), which at the mean year 7 salary level of $3.6 million amounts to a $900,000 salary penalty.

In the third panel of Table 5 I rerun this analysis using a stricter version of the expectations thresholds. Whereas with the original definition a full 30 percent of players played either above or below expectations, by setting the threshold instead at 1.5 times the standard deviation of u_{i_hat}, only 36 players (or roughly 10 percent of the sample) play beyond expectations. This second test finds results comparable to those above. Although the coefficient on the reward for above-expectations play doubles in size when I consider this more select definition, it remains insignificant. In comparison, the penalty rises sharply to 40 percent and is now significant at the one percent level. This hapless group unlucky enough to play especially poorly in year 6 face an on average $1.4 million salary penalty in their free-agency contract.

As a test for whether those who play especially poorly in year 6 also play especially poorly in year 7 – and hence deserve this lower salary – I create similar dummy coefficients on whether year 7 performance was higher than the upper and lower bounds already defined. (Note that even though I am now considering year 7 results I do not recreate the expectations thresholds using year 1 through 6 as the predictors because this will capture the especially poor play in the last year. If I truly believe that this poor year 6 performance is a fluke, then we should consider whether year 7 performance 'recovers'.) Looking at these two dummies – reflecting poorer than expected play in year 6 and year 7 – using the less strict threshold definition I find a correlation of 0.17, and with the stricter version a correlation of 0.11. If we instead simply calculate the correlation of year 6 and 7 performance, this is a much higher 0.68. Based on this, it appears that there is not a high correlation between playing especially badly in year 6 and again in year 7.

Running this same analysis using year 8 as my post-free agency measure I find similar results. In panel two the coefficients on playing beyond expectations are both larger in magnitude but effectively comparable – no significant reward for playing better than expected, and a significant 30 percent penalty for playing poorer than predicted. Panel three finds the same result for playing below expectations – now a 50 percent penalty. (Panel three also shows a negative point estimate for playing above expectations, but the data point is not significant and is based on only 8 players.)

Given these results, it appears that GMs do irrationally overvalue recent information, but only when players play below expectations in the last year of their previous contract. For these, GMs appear to overweight this poor performance in predicting future results, and in the process impose a hefty salary penalty on those players unlucky enough to falter in their last year.

Conclusions

APPENDIX

Appendix Table 1 - Comparison of Post-Free Agency Performance and Salary Coefficients Using On-Base Plus Slugging (OPS) Statistics						
Dependent Variable:	**ln(Performance$_7$)**			**ln(Salary$_7$)**		
	Coefficient	**s.e.**	**p-value[1]**	**Coefficient**	**s.e.**	**p-value[1]**
$\ln(OPS_6)$[2]	0.315	0.074	**0.000**	1.290	0.333	**0.000**
$\ln(OPS_5)$[2]	0.170	0.079	**0.032**	0.663	0.352	0.061
$\ln(OPS_4)$[2]	0.393	0.083	**0.000**	0.899	0.373	**0.017**
Coefficient Ratios:						
(year 6)/(year 5)	**1.85**		-	**1.95**		-
(year 6)/(year 4)	**0.80**		-	**1.43**		-
Age	-0.008	0.005	0.123	-0.072	0.023	**0.002**
Total All Star Appearances (yrs 1-6)	0.001	0.007	0.904	0.104	0.031	**0.001**
Position (omitted = outfield)	-		0.103	-		0.085
1st Base	0.021	0.024	0.368	-0.112	0.106	0.291
2nd Base	-0.058	0.027	**0.031**	0.018	0.120	0.880
3rd Base	0.021	0.031	0.497	0.147	0.138	0.286
Shortstop	-0.014	0.028	0.628	0.308	0.133	**0.022**
Catcher	-0.043	0.028	0.122	-0.133	0.124	0.286
Fielding Percentage (average years 4-6)	0.790	0.794	0.321	8.550	3.555	**0.017**
Team Fixed Effects (year 7)	-		0.216	-		0.239
Team Same (yrs 6 & 7)	-0.009	0.017	0.604	0.354	0.074	**0.000**
Year 7 team located in state of birth	0.026	0.027	0.327	0.068	0.119	0.567
Year 7 Fixed Effects	-		**0.014**	-		**0.000**
Adjusted R^2:	**0.49**			**0.65**		
Sample Size:	**302**			**302**		

NOTES:

(1) For dummy variables (e.g., position or team), the listed value is the joint significance.

(2) $\ln(OPS_t)$ = Log of on-base plus slugging for year t.

Appendix Table 2 - Comparison of Post-Free Agency Performance and Salary Coefficients Using Batting Average (BA)

Dependent Variable:	$\ln(Performance_7)$			$\ln(Salary_7)$		
	Coefficient	**s.e.**	**p-value**[1]	**Coefficient**	**s.e.**	**p-value**[1]
$\ln(OPS_6)^2$	0.201	0.082	**0.015**	1.124	0.399	**0.005**
$\ln(OPS_5)^2$	0.169	0.080	**0.037**	0.537	0.389	0.168
$\ln(OPS_4)^2$	0.238	0.085	**0.006**	0.700	0.413	0.091
Coefficient Ratios:						
(year 6)/(year 5)	**1.19**		-	**2.09**		-
(year 6)/(year 4)	**0.84**		-	**1.61**		-
Age	-0.007	0.005	0.158	-0.093	0.025	**0.000**
Total All Star Appearances (yrs 1-6)	0.005	0.007	0.443	0.164	0.033	**0.000**
Position (omitted = outfield)	-		**0.034**	-		0.061
1st Base	0.006	0.023	0.790	0.012	0.114	0.915
2nd Base	-0.052	0.026	**0.048**	-0.209	0.125	0.096
3rd Base	0.044	0.031	0.158	0.069	0.149	0.642
Shortstop	0.015	0.027	0.573	-0.094	0.130	0.470
Catcher	-0.063	0.028	**0.027**	-0.219	0.136	0.108
Fielding Percentage (average years 4-6)	1.174	0.803	0.145	6.666	3.882	**0.087**
Team Fixed Effects (year 7)	-		0.274	-		0.472
Team Same (yrs 6 & 7)	-0.003	0.017	0.862	0.368	0.081	**0.000**
Year 7 team located in state of birth	0.023	0.027	0.385	0.104	0.129	0.422
Year 7 Fixed Effects	-		0.393	-		**0.000**
Adjusted R^2:	**0.25**			**0.58**		
Sample Size:	**302**			**302**		

NOTES:

(1) For dummy variables (e.g., position or team), the listed value is the joint significance.

(2) $\ln(BA_t)$ = Log of batting average for year t.

Appendix Table 3: Comparison of the Explanatory Power of Alternative Performance Measures

Dependent Variable: ln(Salary7)			
	Coefficient	**(s.e.)**	**p-value**
Slugging On-Base Percentage (SOB)			
Year 6	1.050	0.325	**0.001**
Year 5	0.475	0.347	0.172
Year 4	0.746	0.358	**0.038**
Year 3	0.575	0.335	0.087
Year 2	-0.222	0.301	0.463
Year 1	-0.112	0.288	0.697
Joint Significance (F-Statistic):	9.46		
Adjusted R_2:	0.634		
On-Base Plus Slugging (OPS)			
Year 6	1.337	0.350	**0.000**
Year 5	0.594	0.357	0.098
Year 4	0.942	0.374	**0.012**
Year 3	0.739	0.351	**0.036**
Year 2	-0.243	0.323	0.453
Year 1	-0.092	0.305	0.763
Joint Significance (F-Statistic):	10.64		
Adjusted R_2:	0.643		
Batting Average			
Year 6	1.061	0.414	**0.011**
Year 5	0.443	0.391	0.258
Year 4	0.743	0.413	0.073
Year 3	0.567	0.390	0.147
Year 2	-0.225	0.374	0.548
Year 1	-0.264	0.317	0.405
Joint Significance (F-Statistic):	3.48		
Adjusted R_2:	0.583		